

УДК 004.41: 004.65: 004.67

Л. В. РУДИКОВА, О. Р. МЫСЛИВЕЦ

## О КОНЦЕПЦИИ СОЗДАНИЯ СИСТЕМЫ АГРЕГАЦИИ И ОБРАБОТКИ ДАННЫХ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ

Учреждение образования  
«Гродненский государственный университет имени Янки Купалы»

Актуальным представляется разработка общей концепции и реализация системы хранения и анализа данных практико-ориентированной направленности, одной из подсистем которой является аналитическая система накопления и анализа данных пользователей социальных сетей. Данные, которые пользователи оставляют о себе в социальных сетях, могут быть полезны при решении различных задач. В предлагаемой статье описывается предметная область, связанная со сбором и хранением данных пользователей социальных сетей. Исходя из предметной области, предлагается общая архитектура универсальной системы сбора и хранения данных, которая базируется на клиент-серверной архитектуре. Для серверной части системы приводится фрагмент модели данных, которая связана с накоплением данных из внешних источников. Описывается каркас архитектуры системы. Разрабатываемая универсальная система базируется на информационной технологии складирования данных и для нее характерны следующие аспекты: расширяемая комплексная предметная область, интегрированность хранимых данных, которые поступают из различных источников, инвариантность хранимых данных во времени с обязательными метками, относительно высокая стабильность данных, поиск необходимых компромиссов в избыточности данных, модульность отдельных блоков системы, гибкость и расширяемость архитектуры, высокие требования к безопасности хранимых данных.

Предлагаемая система организует процесс сбора данных и заполнения базы из сторонних источников. Для этого в системе разработан модуль для сбора и преобразования информации из Интернет-источников и отправки их в базу данных. Система предназначена для различных пользователей, заинтересованных в анализе данных пользователей социальных сетей.

**Ключевые слова:** OLTP-система, социальные сети, модель данных, общая архитектура, ETL-процесс; клиент-сервер; подсистема анализа

### Введение

В настоящее время актуальным представляется разработка общей концепции и реализации системы сбора и анализа данных социальных сообществ в сети Интернет, которая может быть рассмотрена в аспекте создания некоего единого хранилища данных, позволяющего производить анализ данных пользователей социальных сетей по необходимым требованиям [1, 2].

Исследование и обобщение предметных областей различной тематики является достаточно актуальным направлением исследований [3, 4]. В настоящее время все большее число исследователей используют большие данные для оценки различных аспектов человеческой деятельности, включая мобильность человека. Например, наборы такого рода данных исполь-

зуются для регионального разграничения в различных масштабах [5], классификации землепользования [6], оптимизации транспортировки [7] и исследований в области транспорта [8], подсчета социально-экономических показателей городских кварталов [9], изучение туристического поведения [10], данных о миграции людей [11, 12] и т. д.

Следует также отметить, что в мире накапливается огромное количество информации, разработаны различные методы, алгоритмы и методологии для анализа данных [см., например, 13, 14, 15], существует достаточное количество необходимого программного обеспечения, технологий и методов для формализации массивов данных, которые позволяют структурировать данные конкретных направлений исследований.



Рис. 1. Упрощенная схема системы сбора и анализа данных социальных сетей

В связи с вышеизложенным, анализ данных социальных сетей является на текущий момент довольно перспективным направлением с широкой сферой применения полученных в процессе анализа результатов. Во-первых, данные социальных профилей пользователей нашли своей применение в области моделирования распространения информационной волны в социальных сетях. Во-вторых, различного рода предвыборные и маркетинговые компании довольно сильно опираются также на данные пользователей социальных сетей: данные такого рода позволяют более эффективно подстраивать агитационные компании под нужды электората. В-третьих, наиболее распространенной сферой применения данных социальных сетей является выявление потребительских предпочтений [16].

Однако на рынке программного обеспечения отсутствуют универсальные программные системы для обработки данных социальных сетей, которые включают в себя получение данных из различных источников, их хранение в определенном формате в течении длительного времени и предоставление расширенного инструментария обработки данных в соответствии с различными группами пользователей. В силу этого предлагаемая концепция системы агрегации данных социальных сетей и, непосредственно, разрабатываемая система является новой, не имеющей аналогов, разработкой, которая будет востребована широким кругом пользователей.

### О предметной области системы агрегации и обработки данных социальных сетей

В рамках рассматриваемой тематики за основу хранимых и обрабатываемых данных принимается информация о пользователях различных социальных сетей (Вконтакте, Facebook, Instagram, LiveJournal и др.), их деятельности, привычках, предпочтениях и взаимодействии с окружающим социумом, как реальным, так и виртуальным. Все данные, которые будут использоваться системой, находятся в открытом доступе и предоставление их пользователями системы является добровольным.

Отметим, что разрабатываемая универсальная система базируется на информационной технологии складирования данных и для нее характерны следующие аспекты: расширяемая комплексная предметная область, интегрированность хранимых данных, которые поступают из различных источников, инвариантность хранимых данных во времени с обязательными метками, относительно высокая стабильность данных, поиск необходимых компромиссов в избыточности данных, модульность отдельных блоков системы, гибкость и расширяемость архитектуры, высокие требования к безопасности хранимых данных.

На рис. 1 представлена упрощенная схема работы системы сбора и анализа данных социальных сообществ в сети Интернет.

Основная концепция предполагаемой системы сбора и анализа данных социальных сообществ в сети Интернет основана на техноло-

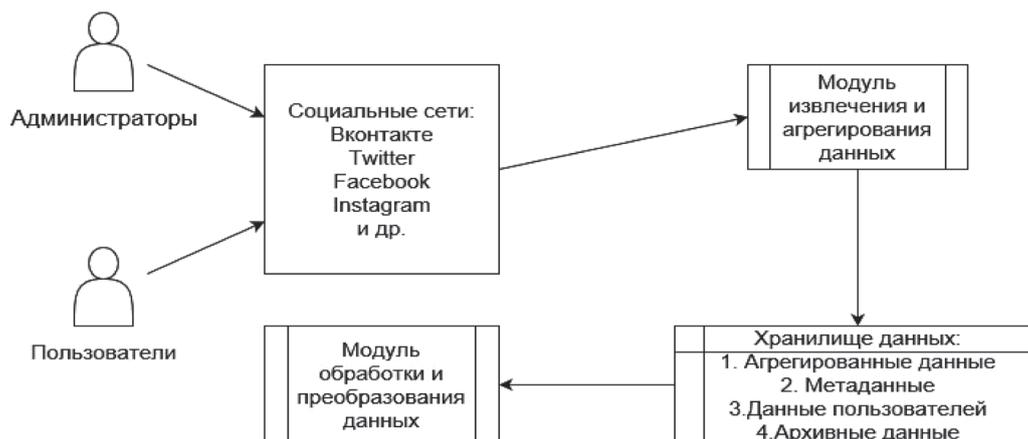


Рис. 2. Общая структура модуля получения и хранения информации

гии складирования данных. Разработка системы ведется с учетом того факта, что функционалом системы будет пользоваться широкий круг лиц для поддержки задач принятия решений предметных областей. В процессе работы возможности системы будут расти, что скажется на ее ресурсоемкости. При построении данной системы следует учитывать принцип модульности, а, конкретно, предполагается наличие как минимум четырех модулей: модуль получения и хранения информации, модуль анализа данных, модуль администрирования системы, модуль, предоставляющий пользователям интерфейс для работы с системой.

#### Общая структура системы хранения и обработки данных социальных сообществ в сети Интернет

Система сбора и анализа данных пользователей социальных сетей состоит из трех основных модулей:

- модуль сбора и хранения информации;
- модуль анализа собранной информации;
- модуль, предоставляющий конечным пользователям системы графический интерфейс.

На рис. 2 представлена общая структура системы сбора и анализа данных социальных сообществ в сети Интернет.

При разработке и проектировании такого рода модуля поднимаются вопросы, связанные с изменениями политик конфиденциальности данных и постепенной модификации структуры и функционала многих социальных сетей.

Отметим, что для разработки основных моделей системы использованы объектно-ориентированный подход и структурная методология построения программных систем.

#### Использование данных социальных сетей

В настоящее время данные, полученные из социальных сетей, используются во многих сферах деятельности, начиная от простого выявления потребительских предпочтений, заканчивая анализом социальных графов, использованием данных пользователей для конкурентной разведки и т. д. Результаты такого анализа могут свидетельствовать о скрытых предпочтениях индивидуума гораздо больше, чем информация в его профиле. Однако данные такого рода представляют собой плохо структурированный набор информации [17]. Это означает, что при разработке системы встает проблема правильной обработки приходящих в подсистему хранения данных. Для решения этой проблемы, как правило, используется ETL-процесс [18], который представляет собой процесс извлечения (получения) информации из OLTP-систем (баз данных), затем ее дальнейшее преобразование к формату хранимых данных в хранилище, и непосредственной загрузки данных в хранилище. Учитывая тот факт, что данные собираются из различных социальных сетей, где информация может быть представлена в различных форматах и нести в себе различную информационную нагрузку, использование имеющегося программного обеспечения не является возможным. Разработка собственного программного решения, которое будет обрабатывать данные согласно необходимым требованиям, является наиболее предпочтительным вариантом. При разработке предполагаемой системы также учитываются возможные ограничения социальных сетей на возможность получения данных. Во всех социальных

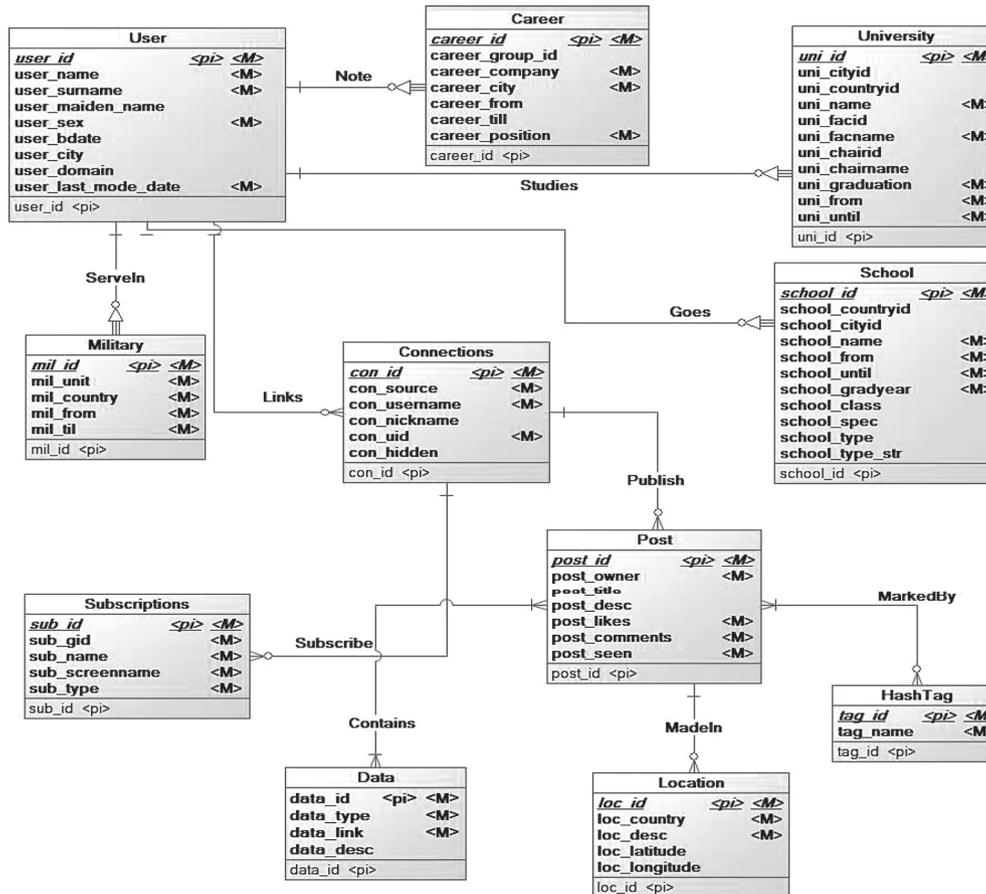


Рис. 3. Фрагмент концептуальной модели базы данных для хранения информации о пользователях

сетях существуют, так называемые, настройки приватности. В зависимости от желания пользователя скрыть некоторые данные о самом себе от определенного круга людей, подсистема может (и получит) только часть данных, что, в последствии, может сказаться на конечном результате анализа его предпочтений. Возможны следующие варианты получения данных.

Во-первых, прямой разбор страницы с профилем пользователя (требует разработки дополнительного функционала, однако он сложен в реализации с учетом растущего количества социальных сетей).

Во-вторых, использование данных, собранных за определенный промежуток времени. Такие данные могут показать динамику изменения предпочтений пользователя.

В разрабатываемой системе агрегации и хранения данных используется второй способ, так как на данный момент ограничения приватности не позволяют получить скрытую информацию как при ручном разборе страницы, так и при использовании соответствующего API.

### Централизованное хранение данных из различных социальных сетей

Как было сказано выше, данные социальных сетей характеризуются плохой структурированностью. В силу этого возникает вопрос, как эффективно хранить данные разных социальных сетей в одной базе данных. Некоторые социальные сети делают упор на обмен сообщениями (Twitter), некоторые на размещение пользователями текстов большого объема (Livejournal), однако большинство социальных сетей представляют возможность агрегации данных различного формата (например, ВКонтакте Facebook). Централизованное хранение данных из различных социальных сетей тоже является важной проблемой, которую необходимо решить при разработке подсистемы сбора и хранения данных.

На рис. 3 представлен фрагмент концептуальной модели базы данных, которая позволит эффективно хранить данные их различных социальных сетей.

Подсистема агрегации и хранения данных на текущий момент включает следующую ин-

формацию о пользователях: основную информацию о пользователе, список социальных сетей в которых пользователь зарегистрирован, информацию об образовании, местах работы. Важным фактором является возможность хранения всех сообществ, на которые пользователь подписан во всех социальных сетях. Отметим, что основной сущностью концептуальной модели является сущность Posts, где хранятся все материалы, опубликованные пользователем в разных форматах (текст, видео, фотографии и т. д.) с возможностью отслеживания хеш-тегов.

### Деперсонализация данных

При разработке подсистемы агрегации и хранения данных важно учитывать и аспект безопасности хранимых данных. Законодательства многих стран очень серьезно относятся к проблеме защиты персональных данных. Согласно требованиям о защите персональных данных, деперсонализация должна обеспечивать не только защиту от несанкционированного использования, но и возможность их обработки [19], т. е. данные после деперсонализации должны обладать рядом свойств, к которым относятся:

- полнота – сохранение информации о конкретном пользователе;
- структурированность – сохранение правильных структурных связей между деперсонализированными данными;
- релевантность – возможность обработки запросов по обработке персональных данных и получения ответов в одинаковой семантической форме;
- применимость – возможность решения задач обработки персональных данных, без предварительного деобезличивания всего объема данных о пользователе;
- анонимность – невозможность однозначной идентификации субъектов данных, полученных в результате деперсонализации, без применения дополнительной информации.

Однако при разработке общей системы анализа социальных сообществ полная деперсонализация не является возможной, т. к. в конечном итоге пользователю системы необходимо знать конкретную информацию о пользователе. Реально возможно лишь частичное достижение анонимности хранимых данных. В системе предлагается шифровать как основную информацию о пользователях (имя, фами-

лия, дата рождения, адрес), так и всю базу данных для предотвращения несанкционированного доступа, в целом. Для предотвращения несанкционированного доступа к самой системе, на которой работает сервис, следует использовать соответствующие политики безопасности, поддерживаемые сервером баз данных, а также некомпьютерные средства контроля.

Данные, собранные подсистемой агрегации и хранения, будут использоваться в следующих задачах.

*Сегментирование аудитории.* В отличие от вышеописанных методов использования данных социальных сетей, разрабатываемая система предполагает возможность сегментирования аудитории по различным признакам. Таким образом, система позволяет разделять пользователей по их интересам и предлагать им соответствующие сообщества. Кроме того, системой делаются предположения о том, какие сообщества заинтересовали бы пользователя в зависимости от публикуемой им информации.

*Для распространения «цепляющего» контента для сообществ.* Продвижение определенного контента, связанного, например, с некоторыми товарами и услугами, будет происходить более эффективно, если предварительно проанализировать возможную аудиторию. Проблема обоюдная: пользователи определенных узконаправленных сообществ могут и не догадываться о том, что в мире появилась вещь или услуга, которая может их заинтересовать и, наоборот, многие производители товаров и услуг охватывают слишком маленькую аудиторию, не пытаясь как-нибудь более эффективно позиционировать себя. Разрабатываемая система представляет информацию о том, на какие сообщества стоит обратить внимания.

Кроме того, общая система анализа социальных сообществ будет использовать полученные данные для построения и анализа графов интересов. Каждый пользователь может получить информацию о людях, с которыми у него могут быть общие интересы в сообществах. Однако возможно использование полученной информации и в противоположных целях: если сообщества являются деструктивными, то данные могут быть использованы для превентивных мер.

### Рекомендации по выбору программных средств

Разрабатываемая система и, в частности, модуль хранения данных, включает в себя единое централизованное хранилище данных, систему обработки данных и систему отчетности. В хранилище данных информация хранится в специально реорганизованном виде в соответствии с разработанной структурой хранения, содержащей необходимые измерения и агрегированные факты предметной области. В связи с этим для разработки модуля сбора и хранения данных используется база данных объектно-реляционного типа MSSQL-Server. В случае использования NoSQL-решений при построении подобных подсистем особых различий в производительности замечено не было [20], однако требования структуризации в NoSQL не выполняются, что затрудняет, в дальнейшем, анализ данных. Программные решения компании Microsoft (СУБД MSSQL, платформы Azure и Business Intelligence) являются более оптимальным решением при построении всей системы, так как дают возможность обеспечить необходимую безопасность данных с использованием механизмов защиты информации, а также обеспечить конфиденци-

альность и целостность данных с использованием Active Directory.

### Заключение

Разработка общей концепции и реализация системы сбора и анализа данных социальных сообществ в сети Интернет является актуальной темой исследований. Социальные сети в настоящее время стали не только местом общения людей, а также и местом, где каждый пользователь собирает и предоставляет различную информацию другим пользователям. Различные взаимосвязи пользователей создают социальные сообщества, анализ и исследование которых позволит определять перспективы использования определенных ресурсов, делать анализ возможных закономерностей по имеющимся массивам данных, а также отслеживать влияние различных факторов на развитие этих сообществ. Предлагаемая система агрегации и обработки данных позволит собирать достаточно обширные сведения по социальным сообществам включая их пользователей, а также получать необходимые аналитические сводки, проводить обработку данных и применять соответствующие методы и алгоритмы Data Mining.

### ЛИТЕРАТУРА

1. Рудикова, Л. В. Об общей архитектуре универсальной системы хранения и обработки данных практико-ориентированной направленности // Л.В. Рудикова / Системный анализ и прикладная информатика. – Мн.: БНТУ, 2017. – № 2. – С. 12–19.
2. Рудикова, Л. В. О моделировании данных предметных-областей практико-ориентированной направленности для универсальной системы складирования и обработки данных // Л. В. Рудикова, Е. В. Жавнерко / Системный анализ и прикладная информатика. – Мн.: БНТУ, 2017. – № 3. – С. 19–26.
3. Belyi, A. Global multi-layer network of human mobility // Alexander Belyi, Iva Bojic, Stanislav Sobolevsky, Izabela Sitko, Bartosz Hawelka, Lada Rudikova, Alexander Kurbatski, Carlo Ratti / International Journal of Geographical Information Science. – 2017. – Volume 31. – P. 1381–1402.
4. Белый, А. Б. Данные сервиса Flickr и структура сообществ стран // А. Б. Белый, Л. В. Рудикова, С. Л. Соболевский, А. Н. Курбацкий / Международный конгресс по информатике: информационные системы и технологии = International Congress on Computer Sciens : Information Systems and Technologies : материалы Междунар. науч. конгресса, Минск, 24 окт.–27 нояб. 2016 г. / БГУ; редкол.: С.В. Абламейко (отв. ред.) [и др.]. – Минск, 2016. – С. 851–855.
5. Amini A. The impact of social segregation on human mobility in developing and industrialized regions / Amini A, Kung K, Kang C, Sobolevsky S, and Ratti C // EPJ Data Science. – 2014. – 3(1):6.
6. Pei T. A new insight into land use classification based on aggregated mobile phone data / Pei T., Sobolevsky S., Ratti C., Shaw S. L., Li T., Zhou, C. // International Journal of Geographical Information Science. – 2014. – 28(9). – P. 1988–2007.
7. Santi P. Quantifying the benefits of vehicle pooling with shareability networks / Santi P., Resta G., Szell M., Sobolevsky S., Strogatz S.H., Ratti C. // Proceedings of the National Academy of Sciences. – 2014. – 111(37). – Pp. 13290–13294.
8. Kung K. Exploring universal patterns in human home/work commuting from mobile phone data / Kung K., Greco K., Sobolevsky S., Ratti C. // PLoS ONE. – 2014. – 9(6):e96180.
9. Hashemian B. Socioeconomic characterization of regions through the lens of individual financial transactions / Hashemian B., Massaro E., Bojic I., Arias J. M., Sobolevsky S., Ratti C. // PloS one – 2017. – 12(11), e0187031.
10. Bojic I. Scaling of foreign attractiveness for countries and states / Bojic I., Belyi, A., Ratt, C., Sobolevsky S. // Applied Geography. – 2016 – 73. – P. 47–52.
11. Sabou M. Visualizing Statistical Linked Knowledge Sources for Decision Support / Sabou M., Hubmann-Haidvogel A., Fischl D., Scharl A. // SemanticWeb. – 2016. – 1. – P. 1–25.

12. **Li Q.** VisTravel: visualizing tourism network opinion from the user generated content / Li Q., Wu Y., Wang S., Lin M., Feng X., Wang H. // *J. Vis.* – 2016. – 19. – P. 489–502.
13. **Рудикова, Л. В.** О разработке системы для поддержки лазерной экспрессной экспертизы. Монография / Л. В. Рудикова – LAP LAMBERT Academic Publishing, 2014. – 134 с.
14. **Барсегян А.** Методы и модели анализа данных: OLAP и DataMining / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб: БХВ-Петербург, 2009. – 336 с.: ил.
15. **Паклин, Н.** Бизнес-аналитика. От данных к знаниям / Н. Паклин, В. Орешков. – СПб.: Питер, 2013. – 704 с.
16. **Информационный центр AfterShock** [Электронный ресурс]. – Режим доступа: [<https://aftershock.news/?q=node/479258&full>]. – Дата доступа: [14.05.2018].
17. **Батура, Т. В.** Методы анализа данных из социальных сетей / Т. В. Батура, Н. С. Копылова, Ф. А. Мурзин, А. В. Проскуряков // *Вестник НГУ. Серия: Информационные технологии.* – 2013. – Т. 11, вып. 3. – С. 5–21.
18. **Среда ETL** (извлечение, преобразование и загрузка) Rational Insight [Электронный ресурс]. – Режим доступа: [[https://www.ibm.com/support/knowledgecenter/ru/SSRL5J\\_1.1.1/com.ibm.rational.raer.overview.doc/topics/c\\_arch\\_etl\\_process.html](https://www.ibm.com/support/knowledgecenter/ru/SSRL5J_1.1.1/com.ibm.rational.raer.overview.doc/topics/c_arch_etl_process.html)]. – Дата доступа: [14.09.2018].
19. **GDPR** — новые правила обработки персональных данных в Европе для международного IT-рынка [Электронный ресурс]. – Режим доступа: [<https://habr.com/company/digitalrightscenter/blog/344064/>]. – Дата доступа: [17.05.2018].
20. **Волушкова В.Л.** Структура данных для хранения информации в социальных сетях / В. Л. Волушкова, А. Ю. Волушкова // *Образовательные ресурсы и технологии* – 2014. – 2(5). – С. 153-157.

## REFERENCES

1. **Rudikova, L. V.** On the general architecture of a universal data storage and processing system of practice-oriented orientation. *Rudikova / System Analysis and Applied Informatics.* – Мн.: BNTU, 2017. – № 2. – P. 12–19.
2. **Rudikova, L. V.** On modeling data of subject-areas of practice-oriented orientation for a universal system of data warehousing and data processing // L. V. Rudikova, E. V. Zhavnerko / *System Analysis and Applied Informatics.* – Мн.: BNTU, 2017. – № 3. – P. 19–26.
3. **Belyi, A.** Global multi-layer network of human mobility // Alexander Belyi, Iva Bojic, Stanislav Sobolevsky, Izabela Sitko, Bartosz Hawelka, Lada Rudikova, Alexander Kurbatski, Carlo Ratti / *International Journal of Geographical Information Science.* – 2017. – Vol. 31. – P. 1381–1402.
4. **Belyi, A. B.** Flickr service data and community structure of countries / A.B. Belyi, L.V. Rudikova, S.L. Sobolevsky, A.N. Kurbatski // *International Congress on Computer Sciens: Information Systems and Technologies: materials of International Scientific Congress, Republic of Belarus, Minsk, 24 October. – 27 Nov. 2016.* / BSU; rare: S.V. Ablameiko (editorial editors) [and others]. – Minsk, 2016. – P. 851–855.
5. **Amini A.** The impact of social segregation on human mobility in developing and industrialized regions / Amini A, Kung K, Kang C, Sobolevsky S, and Ratti C // *EPJ Data Science.* – 2014. – 3(1):6.
6. **Pei T.** A new insight into land use classification based on aggregated mobile phone data / Pei T., Sobolevsky S., Ratti C., Shaw S. L., Li T., Zhou, C. // *International Journal of Geographical Information Science.* – 2014. – 28(9), P. 1988–2007.
7. **Santi P.** Quantifying the benefits of vehicle pooling with shareability networks / Santi P., Resta G., Szell M., Sobolevsky S., Strogatz S.H., Ratti C. // *Proceedings of the National Academy of Sciences.* – 2014. – 111(37). – Pp. 13290–13294.
8. **Kung K.** Exploring universal patterns in human home/work commuting from mobile phone data / Kung K., Greco K., Sobolevsky S., Ratti C. // *PLoS ONE.* – 2014. – 9(6):e96180.
9. **Hashemian B.** Socioeconomic characterization of regions through the lens of individual financial transactions / Hashemian B., Massaro E., Bojic I., Arias J. M., Sobolevsky S., Ratti C. // *PloS one* – 2017 – 12(11), e0187031.
10. **Bojic I.** Scaling of foreign attractiveness for countries and states / Bojic I., Belyi, A., Ratt, C., Sobolevsky S. // *Applied Geography.* – 2016. – 73. – P. 47–52.
11. **Sabou M.** Visualizing Statistical Linked Knowledge Sources for Decision Support / Sabou M., Hubmann-Haidvogel A., Fischl D., Scharl A. // *SemanticWeb.* – 2016 – 1. – P. 1–25.
12. **Li Q.** VisTravel: visualizing tourism network opinion from the user generated content / Li Q., Wu Y., Wang S., Lin M., Feng X., Wang H. // *J. Vis.* – 2016. – 19. – P. 489–502.
13. **Rudikova, L.** About laser express expertise system implementation. Monography / Lada Rudikova. – LAP LAMBERT Academic Publishing, 2014. – 134 p.
14. **Barseghyan, A.** Methods and analysis data models: OLAP and DataMining / A. Barseghyan, M. Kupriyanov, V. Stepanenko, I. Kholod – StP.: BHV-Petersburg, 2009. – 336 p.: il.
15. **Paklin, N.** Business-analytics: from data to knowledge / N. Paklin, V. Oreshkov. – StP.: Piter, 2009. – 624 p.
16. **Информационный центр AfterShock** [Электронный ресурс]. – Режим доступа: [<https://aftershock.news/?q=node/479258&full>]. – Дата доступа: [14.05.2018].
17. **Batura, T. V.** Methods of analyzing data from social networks / T. B. Batura, N. S. Kopylova, F. A. Murzin, A. V. Proskuryakov // *Bulletin of NSU. Series: Information technology.* – 2013. – Vol. 11, issue. 3. – P. 5–21.
18. **ETL environment** (extraction, transformation and loading) Rational Insight [Electronic resource]. – Access mode: [[https://www.ibm.com/support/knowledgecenter/ru/SSRL5J\\_1.1.1/com.ibm.rational.raer.overview.doc/topics/c\\_arch\\_etl\\_process.html](https://www.ibm.com/support/knowledgecenter/ru/SSRL5J_1.1.1/com.ibm.rational.raer.overview.doc/topics/c_arch_etl_process.html)]. – Access date: [14.09.2018].
19. **GDPR** – new rules for the processing of personal data in Europe for the international IT market [Electronic resource]. – Access mode: [<https://habr.com/company/digitalrightscenter/blog/344064/>]. – Access date: [May 14.05.2018].

20. Volushkova V. L. Data structure for storing information in social networks / V. L. Volushkova, A. Y. Volushkova // Educational resources and technologies. – 2014. – 2 (5). – P. 153–157.

Поступила  
23.05.2018

После доработки  
25.10.2018

Принята к печати  
30.11.2018

Rudikova L. V., Myslivec O. R.

## ABOUT A CONCEPT OF CREATING A SOCIAL NETWORK USERS INFORMATION AGGREGATION AND DATA PROCESSING SYSTEM

Yanka Kupala State University of Grodno

*The development of a general concept and implementation of a data-storage and analysis system for practice oriented data, one of the subsystems of which is an analytical system for the accumulation and analysis of data from users of social networks, is topical. The development of a general concept and implementation of a data-storage and analysis system for practice oriented data, one of the subsystems of which is an analytical system for the accumulation and analysis of data from users of social networks, is topical. Data that users leave about themselves in social networks can be useful in solving various tasks. The proposed article describes the subject area associated with the collection and storage of data from users of social networks. Proceeding from the subject area, the general architecture of the universal data collection and storage system is proposed, which is based on the client-server architecture. For the server side of the system, a fragment of the data model is provided, which is associated with the accumulation of data from external sources. The framework of the system architecture is described. The developed universal system is based on the information technology of data warehousing, and it has the following aspects: an expandable complex subject area, the integration of stored data that come from various sources, the invariance of stored data in time with mandatory labels, relatively high data stability, the search for necessary trade-offs in data redundancy, modularity of individual system units, flexibility and extensibility of the architecture, high security requirements vulnerable data.*

*The proposed system organizes the process of collecting data and filling the database from external sources. To do this, the system has a module for collecting and converting information from third-party Internet sources and sending them to the database. The system is intended for various users interested in analyzing data of users of social networks.*

**Keywords:** OLTP-system, social networks, data model, general architecture, client-server; analysis subsystem, ETL-process.



### Рудикова Лада Владимировна

Ул. Ожешко. 22, 212, Гродно, 230023, Беларусь, rudikowa@gmail.com

Тел. +375 297 816 355; ГрГУ, зав. кафедрой современных технологий программирования, кандидат физико-математических наук

**Lada Rudikova** is the Head of Modern Programming Technologies Department of Yanka Kupala State University of Grodno (YKSUG). Ph.D. degree in physical and math.

The main line of her scientific researches – management theory, information systems design, databases, CASE, data mining, business intelligence. She actively participates in international conferences. She is the author of more than 280 scientific works and books related to computer technology and data processing, a technical writer of the publishing house «BHV-St Petersburg».



### Мысливец Олег Романович

Ул. Ожешко. 22, 212, Гродно, 230023, Беларусь, myslivec.oleg@yandex.ru

Тел. +375 336 882 079, ГрГУ, преподаватель кафедры современных технологий программирования, магистр технических наук

**Oleg Myslivec** is a lecturer in Yanka Kupala Grodno State University. Master degree in computer science.

The main line of his scientific researches – information systems design, databases, CASE, big data analysis.

*Результаты работы получены в процессе выполнения ГПНИ «Разработка методологии и средств построения универсальных систем хранения, обработки и анализа структурированных данных большого объема практико-ориентированной направленности» (№ гос. регистрации 20162266).*